YOLO 모델을 이용한 동영상 하이라이트 추출 프로그램

**컴퓨터공학과 2017103949 고범석**

**컴퓨터공학과 2016104112 김영빈**

**컴퓨터공학과 2017103975 김용재**

**컴퓨터공학과 2017104013 이윤종**

**요약**

코로나 사태가 회복되어 가면서 여행객의 수요가 늘었지만, 여행의 결과물인 사진 및 동영상, 특히, 동영상에 대한 관리 기능이 부족한 상황이다. 동영상은 여행 중의 즐거운 순간들을 생동감 있게 표현해줄 수 있는 매체인데, 재생 시간이 너무 긴 경우에는 다 보지 못한다는 단점도 가지고 있다. 이에 본 연구에서는 동영상의 내용 중 의미 있는 순간들을 기록한 영상 하이라이트를 제작하여 사용자가 편하게 관리 혹은 감상할 수 있는 애플리케이션을 제안한다.

1. **서론**
   1. **연구 배경**

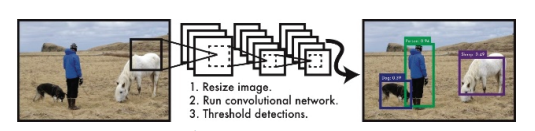
최근 코로나 사태가 회복됨에 따라 여행을 즐기지 못한 사람들의 욕구가 상승하면서, 여행에 대한 수요가 증가하고 있다. 또한, 자연스럽게 여행지에서의 추억을 위한 사진과 동영상 또한 증가한다. 하지만, 이렇게 많이 발생한 사진과 동영상 중 특히 동영상은 길이가 길고 연속적이다 보니 관리하기 까다롭다. 이 문제를 해결하기 위해 본 연구에서는 객체 인식 모델 및 표정 인식 모델을 이용하여 긴 영상에서 짤막한 하이라이트 조각들을 만들어 사용자에게 제공해주는 애플리케이션을 제안한다.

* 1. **연구 목표**

본 연구의 목표는 영상에 등장하는 인물들의 행복한 순간을 포착하여 이를 하이라이트 영상으로 남겨줌으로써 보다 손쉽게 추억을 관리하는 것이다. 이를 위해 분석하고자 하는 영상에서 사람들이 웃음을 짓는 시간대를 추출한다. 첫번째 목표로 클라이언트에서 촬영된 영상을 압축한다. 두번째 목표로 압축된 영상을 클라우드 서버로 전송한다. 세번째 목표로 서버는 클라우드 속 영상을 YOLO 모델 및 표정 인식 라이브러리로 분석하여 요구사항에 맞는 장면을 찾아 해당 시간대를 클라이언트에 전송한다. 최종 목표로 전달 받은 시간대를 바탕으로 클라이언트에서 해당 영상을 잘라 하이라이트 영상을 만들어 저장한다.

1. **관련 연구**
   1. **객체 인식 기술**
      1. **YOLO**

이미지 내에 존재하는 객체와 해당 객체의 위치는 이미지를 한번 만 보고 예측할 수 있는 알고리즘이다. 이는 분류 대상의 객체로 감지하는 대신에, 경계 박스를 다차원적으로 분리하고 클래스 확률(Class Probability)을 적용하여 하나의 회귀 문제로 접근한다. 입력 이미지는 CNN을 통해 텐서(Tensor)의 그리드 형태로 나누고, 각 구간에 따라 객체인 경계 박스와 클래스 확률을 생성하여 해당 구역의 객체를 인식한다



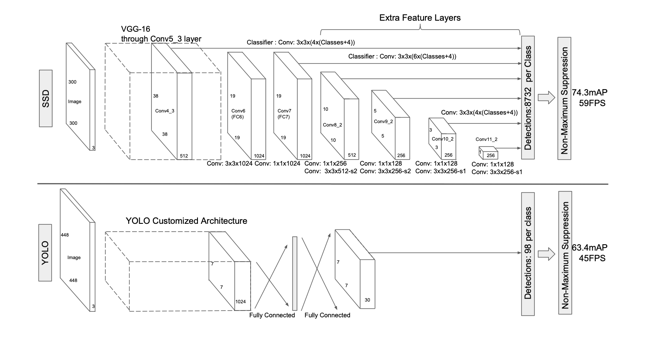
* + 1. **R-CNN**

R-CNN은 설정한 Region을 CNN의 feature로 활용하여 객체 인식을 수행하는 신경망이라는 의미를 담고 있다. 이 모델은 selective search을 통해 2,000개의 region proposal을 먼저 뽑아낸 후, CNN모델에서는 selective search을 통해 찾아낸 2,000개의 영역을 227x227 크기로 조정하고 CNN모델을 통과하여 4096 크기의 특징 벡터를 뽑는다. CNN을 통해 추출한 벡터를 가지고 각각의 클래스 별로 SVM classification을 학습시키고 객체 인식을 수행한다. 위 과정 속 selective search을 통해 찾은 박스의 위치는 정확하지 않아 이런 localization error을 줄이기 위해 CNN feature을 이용하여 bounding box regression 모델을 사용한다. R-CNN은 2,000개의 region proposal을 뽑고 각 영역마다 CNN을 수행하기 때문에 수행시간이 느리다.

Diagram

Description automatically generated

* + 1. **SSD**

SSD는 YOLO와 달리 컨볼루션 과정을 거치는 중간 중간의 피쳐맵들에서 Object Detection을 수행한다. 이 모델은 VGG16을 base network로 사용하고, 보조 network를 추가한 구조를 가지고 있다. SSD는 이미지를 입력받아서 이미지 넷으로 사전 학습된 VGG의 Conv5\_3 layer까지 통과하여 피쳐를 추출한다. 추출된 피쳐맵을 컨볼루션을 거쳐 그 다음층에 넘겨주는 동시에 Object Detection을 수행한다. VGG를 통과하여 얻은 피쳐맵을 대상으로 컨볼루션을 진행하여 최종적으로 1x1 크기의 피쳐맵까지 뽑는다. 각 단계별로 추출된 피쳐맵은 Detector & Classifier를 통과시켜 Object Detection을 수행한다.

* 1. **표정 인식 기술**

표정인식 기술은 단일 이미지나 이미지 시퀀스에서 얼굴의 표정을 인식하는 기술이다. 과거에는 SIFT, HOG, LPQ 등의 알고리즘을 통해서 얼굴의 특징을 추출하여 분류기를 통해 얼굴 표정 인식을 수행했었다. 최근에는 딥러닝 기술의 발전으로 인해 위의 알고리즘 대신 CNN과 RNN을 활용한 얼굴 표정 인식 기술들이 연구되고 있다.

* + 1. **HaarCascade**

**스크린샷이(가) 표시된 사진

자동 생성된 설명**하르 캐스캐이드는 객체 인식 기술 중의 하나로 얼굴 표정 분석 기능을 제공한다. 이는 사각형의 필터 집합을 사용하여 사람 얼굴의 특징이 있는지 판단한다. 또한, 캐스케이드 분류기를 통해서 각 단계 별로 비교할 얼굴의 특징을 두고 얼굴이라면 다음 단계로 진행하여 최종적으로 얼굴을 인식해낸다. 캐스케이드 분류기는 오브젝트가 포함된 이미지와 오브젝트가 없는 이미지를 사용해서 학습시킨다. 픽셀 단위로 연산하지 않고 직사각형 영역으로 구성되는 특징을 사용해서 동작 속도가 빠르다는 장점이 있다.

1. **프로젝트 내용**
   1. **시나리오**
2. **동영상 파일 업로드**

영상은 휴대폰 기본 카메라 앱을 사용하여 찍고 해당 영상을 애플리케이션(클라이언트) 내의 업로드 버튼을 통해 업로드 한다.

1. **동영상 압축 및 클라우드 서버 업로드**

애플리케이션에서는 업로드 된 동영상을 MPEG-4 형식으로 압축하여 클라우드에 업로드 한다.

1. **동영상 하이라이트 시간대 추출**

서버는 클라우드로부터 압축된 영상 데이터를 받아 객체 인식 및 표정 인식 모델을 사용하여 웃고 있는 사람이 나오는 시간대를 추출하고 이를 클라이언트에게 반환한다.

1. **하이라이트 제작**

클라이언트는 서버에서 받아온 하이라이트의 시간대를 받아 원본 영상에서 시간대 별로 추출하여 하이라이트 토막을 제작하고 사용자에게 각각의 토막을 동영상 형태로 제공한다.

* 1. **요구사항**
     1. **동영상 압축 및 클라우드 서버 전송에 관한 요구사항**

- 애플리케이션에서는 처리할 동영상을 MPEG-4 형식으로 압축한다. 동영상의 크기가 매우 클 것으로 판단하기 때문에 클라우드 서버 전송 시 많은 시간이 소요될 것으로 예상되어 이를 해결하고자 동영상 추가 압축을 수행한다.

- 압축된 영상을 클라우드에 업로드한다. 클라이언트에서 압축된 동영상을 서버로 바로 전송할 경우 커넥션을 계속 유지해야 하는 단점이 있기 때문에 S3 서버에 저장해두고 서버가 이를 가져와서 영상 처리를 수행한다.

* + 1. **하이라이트 추출에 관한 요구사항**

- 영상 내 하이라이트를 가장 효과적으로 판별, 추출할 수 있는 방법을 고안한다.

1. 압축된 영상 내 사람들의 얼굴 인식

2. 얼굴이 인식된 사람들의 웃음 인식 및 판별

3. 웃었다고 판별되기 시점의 2초전 시점과 판별이 끝난 시점의 2초후 시점을 저장 후 클라이언트에게 보내준다.

- 이 요구사항들을 처리하기 위해 필요한 것은 영상처리이다. 영상 내 사람들의 얼굴 인식을 구현하기 위해 오픈소스 영상 분석 모델인 Yolo을 활용한다.

- 사람들의 얼굴을 인식한 순간 사람들이 웃고 있는지를 판별하기 위해 HaarCascade를 사용한다.

- 사람들이 모두 웃고 있다고 판별되면 판별된 시점의 2초전 시간과 판별이 끝난 시점의 2초 후 시간을 저장해 두고 모든 영상 분석이 끝나면 클라이언트에게 저장해둔 시간 정보를 보내준다.

* + 1. **하이라이트 결과에 대한 요구사항**

- 동영상을 각 시간 데이터로 잘라서 S3에 저장하도록 구현한다.

- 클라이언트가 S3를 통해 각 동영상들을 볼 수 있도록 구현한다.

- 각 동영상에 대해 삭제가 가능하도록 구현한다.

1. **향후 일정 및 역할 분담**

|  |  |
| --- | --- |
| ~ 4.9 | 관련 연구 조사 및 객체 인식, 표정 인식모델 조사 및 학습 |
| 4.10 ~ 4.30 | 영상 압축 방식 조사, 선택 및 API 구현 |
| 5.1 ~ 5.10 | 클라우드 서버 구축 및 클라이언트와의 연동 테스트 |
| 5.11 ~ 5.24 | 하이라이트 시간대 추출 기능 구현 |
| 5.25 ~ 6.08 | 하이라이트 영상 제작 및 업로드 구현 |

|  |  |
| --- | --- |
| 공통 | 관련 모델 조사(YOLO, OpenCV, HaarCascade) |
| 김영빈 | 영상 하이라이트 제작 알고리즘 개발, HaarCascade 적용 |
| 김용재 | 학습 데이터 수집 및 YOLO 모델 학습 |
| 고범석 | 영상 하이라이트 제작 알고리즘 개발, YOLO 적용 |
| 이윤종 | Swift 클라이언트 개발 |

1. **결론 및 기대효과**

코로나 사태가 안정되어 가면서 많아진 여행객들의 추억 저장을 위하여 동영상의 하이라이트 제작 애플리케이션을 제안했다. 본 애플리케이션에서는 사용자의 동영상에서 추억이 될 수 있는 부분을 파악하고 해당 부분들을 각각의 영상으로 만들어 날짜 별 폴더에 저장하고 언제든지 사용자가 꺼내볼 수 있는 기능을 제공한다. 이를 통해, 언제 어떤 사람들과 여행을 다녀왔는지, 어떤 즐거운 순간들이 있었는지 손쉽게 추억할 수 있고 여행의 즐거움을 여행하면서, 여행에 다녀와서 두 번 즐길 수 있게 해준다. 또한, 최종 하이라이트 영상에서 마음에 들지 않는 부분은 삭제할 수 있기 때문에 효율적인 추억 관리 기능을 제공한다. 영상 전송 단계에서는 영상을 압축하고 하이라이트 추출 단계에서는 하이라이트의 시간대를 활용하기 때문에 빠른 처리 시간으로 원본 영상의 화질을 유지하여 하이라이트를 볼 수 있는 장점이 있다. 이 애플리케이션을 사용하여 사용자만의 추억 영상 만들기와SNS 사용량이 정점에 달하는 요즘 편한 하이라이트 영상 공유를 기대한다.

1. **참고 자료**
2. 이용환, 김영섭. 객체 검출을 위한 CNN과 YOLO 성능 비교 실험. 반도체디스플레이기술학회지. 2020; 19(1) 85-92.
3. 이민규, 송병철. (2019). 얼굴표정 인식 기법의 최신 연구 동향. 한국미디어공학회 학술발표대회 논문집 , 126-128.
4. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* , 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.
5. Liu, W. *et al.* (2016). SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0\_2
6. Girshick, Ross & Donahue, Jeff & Darrell, Trevor & Malik, Jitendra. (2013). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 10.1109/CVPR.2014.81.
7. Uijlings, Jasper & Sande, K. & Gevers, T. & Smeulders, A.W.M.. (2013). Selective Search for Object Recognition. International Journal of Computer Vision. 104. 154-171. 10.1007/s11263-013-0620-5.